분산 유전자 알고리즘을 이용한 컬러 이미지의 영역분할

조찬윤[†] · 김상균[††]

요 약

컬러 영상들은 응용 분야별로 독특한 특성을 가지고 있다. 따라서 실용적인 영상 분할 시스템을 구축하기 위해서는 특정 영상에 독립적인 방법을 개발할 필요가 있다. 본 논문에서는 분산 유전자 알고리즘을 이용한 컬러 유방암조직영상의 분할 방법을 제안한다. 컬러 유방암조직영상에서 양성세포핵과 음성세포핵을 분할하기 위해서, 컬러 정보를 효과적으로 반영하는 개선된 평가함수 및 유전연산 기반의 분산 유전자 알고리즘을 이용한다. 또한 성능 향상을 위하여 영상에서의 대표색을 추출하여 초기치로 활용한다. 유효성을 입증하기 위한 실험에서 안정된 분할 결과를 보였으며, 이는 제안한 방법이 제한된 컬러를 가진 제한된 개체를 분할할 때 실용화 할 수 있음을 제시한다.

Region Segmentation of a Color Image using a Distributed Genetic Algorithm

Cho Chan-Yun[†] and Kim Sang-Kyoon[††]

ABSTRACT

Color images from various application areas have their own characteristics. Practical segmentation systems need specialized methods to deal with the characteristics. In this paper, we propose a distributed genetic algorithm based segmentation method for color breast carcinoma cell images. To extract positive nuclei and negative nuclei from the cell images, a distributed genetic algorithm with improved genetic operations and an evaluation function is used. As initial values, representative colors from images are introduced to work well with the cell images. A test to verify the validity of the proposed method shows well-segmented images. This result suggests that the method is pertinent to be put into practical use for the images having limited objects with limited colors.

1. 서 론

컴퓨팅 환경의 발달로 인하여 자동적인 컬러 영상 분석은 다양한 분야에서 활발히 연구되고 있다.

특히 의료영상분야에서는 점차적으로 영상에 대한 분석이 컴퓨터에 의해 수행되고 있는 추세이다 [1].아직까지는 각종 의료영상에 대해 의사가 직접 육안으로 판단하는데 많이 의존하고 있지만, 정확한 진단을 돕기 위해서는 의료영상에 대한 자동적인 분석이 요구된다. 이러한 영상분석에서 영역분할은 전단계이자 핵심적인 기술이다. 따라서 영상 정보 시스템의 구축을 위해선 실용적인 영역분할 방법에 대한 연구가 필수적이다.

제시되고 있는 영역분할 방법은 경계선 기반 방법(boundary based method)과 영역 기반 방법(region based method)으로 크게 구분할 수 있다. 경계선 기반 방법은 각 화소의 위치에서 그래디언트를 사용하여 에지를 검출한 후에 그래프 탐색, 동적 프로그래밍(dynamic programming), 하프 변환 등의 방법을 사용하여 인접한 에지를 연결하는 방법을 사용한다 [2,3]. 영역 기반방법으로는 영상 내의 히스토그램을 사용한 임계치 방법, 신경망을 이용한 화소 분류방

본 논문의 연구는 1998년도 인제대학교 학술연구조성비 보조에 의해 수행되었음.

[†] 인제대학교 전산학과 석사과정
[††] 정회원, 인제대학교 정보컴퓨터공학부 조교수

법, 영역 내의 유사성 및 연결성을 이용하여 영역을 확장시켜 가는 영역 성장(region growing)기법 등이 있다[4-6]. 영역간의 유사성을 추정하기 위한 통계적 모델로서는 MRF(Markov Random Field) 모델을 들 수 있으며, 모수 추정을 위한 도구로서는 시뮬레이티드 어닐링과 유전자 알고리즘을 이용한다[7-9]. 특히 유전자 알고리즘은 계산상의 제약을 받지 않는 영역 기반의 확률적인 탐색 기법으로 여러 분야에서 최적화 도구로 이용되고 있다.

효과적인 영상분할을 위하여 위와 같은 다양한 방법들이 제시되고 있지만, 주로 명암값(gray-level)으로 표현된 영상을 대상으로 하고 있다. 특히 유전자 알고리즘의 특성을 수용하면서 자율적인 영상 분할을 위해서 제시된 분산 유전자 알고리즘은 좋은 성능을 보였으나 명암값 영상이 실험대상이었다[10]. 실생활의 영상을 반영하기 위하여 컬러 정보를 이용하는 방법에서도, 컬러 영상들은 분야별로 독특한 특성을 가지고 있기 때문에, 제안된 대부분의 영상 분할 방법은 다양한 종류의 영상에 일반적으로 적용하기 힘들다. 따라서 실용적인 영상 분할 시스템을 구축하기 위해서는 일반적인 분할 방법보다는 특정 영상에 독립적인 방법을 개발할 필요가 있다.

본 논문에서는 분산 유전자 알고리즘을 이용한 컬러 유방암조직영상의 분할 방법을 제안한다. 컬러 유방암조직영상은 암의 진행 정도를 파악하고 진단하기 위해 사용되는 영상이다. 컬러가 아닌 흑백영상이지만, 특수한 약품을 이용하여 진단자가 관심 있는 영역이 특별한 컬러를 가지도록 처리한 영상이다. 영상분석 시 고려 대상은, 암의 진행정도를 결정하는 조직내의 양성세포핵과 음성세포핵의 분포정도, 형태학적 양상 등이다. 이러한 컬러 유방암조직영상을 분할하기 위하여, 컬러 정보를 효과적으로 반영하기 위한 개선된 평가함수 및 유전연산 기반의 분산 유전자 알고리즘을 이용한다. 제시하는 방법은 제한된 개체들을 영상에서 분할해내는 방법으로 효과적이다. 즉 유전 연산이 전체 집단에서 이루어지는 일반적인 유전자 알고리즘과는 달리, 분산 유전자 알고리즘은 지역적으로 분산된 부분모집단에서 이루어지므로, 관심 있는 개체에 대한 영역 기반의 탐색을 수행할 수 있다. 또한 유방암조직영상은 특수한 컬러 처리를 한 영상이므로 관심 있는 개체들은 각각 독특한 컬러를 가지게 되고 컬러의 수도 자연 영상에 비하여 훨씬 적다. 따라서 모수 추정을 위한 최적화 도구들이 가지는 근본적인 문제인 초기화 문제를 간단한 전처리로 해결하여 성능을 향상시킬 수 있다. 더불어 분산 유전자 알고리즘의 특성인 지역적인 국부집단에 대한 유전 연산은, 병렬 수행을 가능하게 한다. 그러므로 본 논문에서 제시하는 유방암조직영상의 분할 방법은 결과의 질적 수준과 수행 속도 문제를 동시에 해결함으로써 실용화할 수 있는 요건을 갖추고 있다.

2. 유전자 알고리즘

유전자 알고리즘은 다윈의 진화론에 바탕을 둔 문제 해결 방법이다. 이 알고리즘의 목표는 잠재적인 해 집단의 반복적인 변형을 통해 최적 혹은 최적에 가까운 해를 얻고자 하는 것이다. 어떤 문제에 대해 가능한 해 집단은 비트 열로 표현되며 이를 염색체(chromosome)라 한다. 염색체 수의 크기는 문제에 따라 적합하게 주어져야 한다. 이러한 염색체 집단은 적합도 함수에 근거한 유전 연산자에의 재생연산을 거쳐 세대를 반복하며 개선된다.

일반적인 유전자 알고리즘은 임의의 값으로 초기화된 개체들의 집합으로 시작한다. 각각의 개체는 상대적인 문제해결 능력에 따라 그 적합도(fitness)가 평가되며, 적합도에 따라 다음 세대(generation)에 부모의 유전자가 복제(reproduce)되는 정도를 달리 함으로써 우성 형질을 지닌 개체들은 열성 형질을 지닌 개체들에 비하여 더욱 많은 자식을 생성할 수 있도록 유도된다. 이러한 선택(selection)메카니즘은 다윈의 진화론에서 적자 생존의 원리(survival of the fitness)에 연유한다.

이 알고리즘에서 사용하는 연산자는 재구성 제어기(reproduction operator), 교차기(crossover operator), 돌연변이기(mutation operator)가 있다. 재구성 제어기는 적합도에 따라 모집단에서 두 개의 염색체를 선택하는 기능을 담당한다. 교차기는 재구성 제어기에 의해 선택된 두 염색체의 인자 값을 서로 맞바꾸어 새로운 염색체를 생성한다. 이때 무작위로 선택된 교차위치와 개수에 따라 여러 제어기로 나누어진다. 또한, 인자들 간의 상호 관련성, 즉 인자들 간의 순서, 인자들 간의 인접성 등에 따라 교차 알고리즘이 달라진다. 돌연변이기는염색체 내의 인자를 무작위로 선택하고, 그 값을 임의로 바꾸어 새로운 염색체를 만

드는 기능을 한다. 그러므로 모집단의 특성을 다양화 시키며 국부최적을 벗어나게 하는 역할을 담당한다. 표 1은 기본적인 유전자 알고리즘이며 표 2는 유전 연산의 예이다. 교차는 두 부모 *P1*과 *P2*상의 교차점을 기준으로 각각 한 부분씩을 상속받아 자손 *O1*과 *O2*를 생성한다.

표 1. 기본적인 유전자 알고리즘

```
Simple Genetic Algorithm()
{
    initialize(Population);
    do {
        evaluation(Population);
        Selection_Set = selection(Population);
            // reproduction operator
        recombine(Selection_Set);
            // crossover operator, mutation operator
    }while(stop criterion has not been met);
}
```

표 2. 유전 연산자(교차와 돌연변이)

연산자	부 모	자 손
교차	*P1*: 1 1 0 | 1 0 1 0	*O1*: 1 0 1 1 0 1 0
	P2: 1 0 1 | 0 0 1 1	*O2*: 1 1 0 0 0 1 1
돌연변이	1 0 *1* 0 0 1 1	1 0 ***0*** 0 0 1 1

유전자 알고리즘이 기존의 탐색 또는 최적화기법과 다른 특징은 다음과 같다: 1)유전자 알고리즘은 점(point)이 아닌 군(population)에 기반한 탐색 방법이다; 2)유전자 알고리즘에서의 탐색은 확률적 연산자를 사용하여 수행된다; 3)탐색공간에 대해 연속성이나 미분가능성 등의 제약을 요구하지 않는다. 이와 같이 유전자 알고리즘은 그 개념과 이론이 단순하고, 해의 탐색 성능이 우수하여 여러 분야의 다양한 최적화 문제에 적용되고 있다.

3. 컬러 유방암조직영상

컬러 유방암조직영상은 암의 진행 정도를 파악하고 진단하기 위해 사용되는 영상이다. 실제 영상은 흑백 영상이나 세포를 특별한 약품으로 처리하여 세포핵들과 배경이 특정한 컬러를 띠도록 한 영상이다. 영상의 획득 및 진단 과정은 다음과 같다. 유방암 환자로부터 채취한 세포핵 조직 절편을 면역조직화학적으로(IHC) 염색한다. 그리고 CCD 칼라 비디오 카메라가 연결되어 있는 Olympus 광학 현미경을 통해 40배 확대된 영상을 획득한다. 진단자는 이 영상에서 양성세포핵과 음성세포핵의 분포정도와 형태학적 양상 등을 근거로 암의 진행 정도를 결정하고 적절한 치료 계획을 세운다[4].

유방암조직영상은 진단자가 관심을 가지는 부분이 특별한 컬러를 띠도록 처리하였기 때문에 컬러의 종류가 다양하지 않다. 갈색은 양성세포핵이고, 검회색은 음성세포핵이며, 주황색과 노란색은 배경이다. 그리고 양성세포가 변하여 음성세포가 되는 것이기 때문에 양성세포에서 음성세포로 가는 중간 과정의 세포는 갈색과 검회색의 중간색을 띄고 있다. 그림 1은 컬러 유방암조직영상의 예이다. 이러한 유방암조직영상은 제한된 색을 가진 제한된 개체들로 이루어진 영상이다.

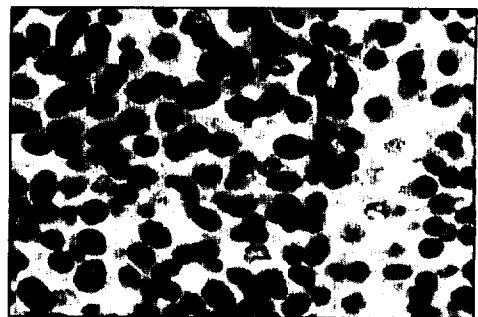



그림 1. 컬러 유방암조직영상

4. 분산 유전자 알고리즘을 이용한 영상의 영역분할

유전자 알고리즘을 실제 응용하기 위해 해결해야 하는 문제는 먼저 문제의 특징들을 어떻게 염색체라는 데이터 구조로 전환하는가에 달려 있다. 즉 영상의 특징에 따라 염색체 내의 인자의 종류, 개수, 그리고 그 값의 범위를 결정해야 한다. 그리고 교차 염색체의 선택과 교차방법, 돌연변이 방법 등이 주어진 문제에 따라 효과적으로 달라져야 한다. 또한, 최적의 해를 판단하는 기준을 정하고 이를 목적함수로

표현해야 한다. 이러한 방법론들은 기존의 명암값 영상과는 달리, 컬러 영상을 효과적으로 표현할 수 있는 구조로 개선되어야 한다. 부가적으로 탐색 과정에서 초기화는 성능에 커다란 영향을 미친다. 즉 임의의 값보다는 영상에서 의미 있는 값을 찾아 최초 모집단을 초기화함으로써 해의 질적 향상을 꾀할 수 있다. 또한 이러한 일반화된 초기화 과정은 의료영상의 분석에서 중요한 요소인 재현가능(Reproducibility) 해야함을 만족시킨다. 다음은 제안하는 방법의 단계별 기술이다.

4.1 염색체

본 논문에서 제시하는 염색체는 실제 영상의 각 화소의 특성벡터에 대응하는 가상벡터이다. 실제 컬러 영상의 각 화소의 특성벡터는 원 영상의 RGB 색공간이다. 따라서 이에 상응하는 염색체의 값도 R, G, B 각각 0에서 255 사이의 값이므로 각각 8 비트씩 24비트의 비트열로 표현된다.

염색체와 마찬가지로 레이블도 원 영상의 각 화소마다 배정된다. 이러한 레이블은 영상분할의 결과를 나타내는데, 결과적으로 이웃하는 화소가 같은 레이블을 가지면 같은 영역에 속함을 의미한다. 초기화 시 동일한 레이블을 가지는 화소에 대한 염색체는 같은 종류의 초기값을 가지게 된다. 따라서 초기 염색체의 종류는 레이블 개수와 동일하다. 그림 2는 염색체와 레이블을 가지는 가상 벡터의 예이다.

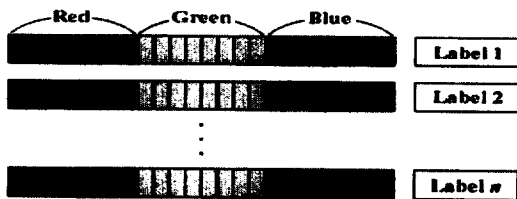



그림 2. 가상 벡터

염색체는 평가, 선택, 유전연산(교차와 돌연변이)의 세가지 단계를 거쳐 새로운 세대로 재형성된다. 이 때 각 화소마다 프로세서가 배정되어 국부적이고 독립적으로 진화할 수 있다. 이러한 방식은 메쉬(mesh) 구조의 병렬 컴퓨터에서 효과적으로 수행 될 수 있다.

4.2 초기화

일반적으로 염색체의 초기값 지정 시 임의의 값을 이용한다. 본 논문에서는 각 영역의 컬러를 대표하는 색들을 추출하여 초기값으로 활용한다.

표 3. 대표값 추출 방법

	추출 방법
영역당 1개 추출	*Init1* = *Mean*
영역당 3개씩 추출	*Init1* = *Min* + (*Mean* - *Min*) / 3 *Init2* = *Mean* *Init3* = *Max* - (*Max* - *Mean*) / 3
영역당 5개씩 추출	*Init1* = *Min* + (*Mean* - *Min*) / 3 *Init2* = *Init1* + (*Mean* - *Init1*) / 2 *Init3* = *Mean* *Init4* = *Max* - (*Max* - *Mean*) / 3 *Init5* = *Init4* - (*Init4* - *mean*) / 2

수작업으로 양성세포핵에 대한 갈색영역, 음성세포핵에 대한 검회색영역 그리고 배경에 해당하는 주황색과 노란색영역에서 각 20개씩의 표본을 추출한다. 추출된 표본들의 RGB값 각각의 평균(*Mean*), 최대(*Max*), 최소값(*Min*)을 구한다. 이러한 값들을 근거로 영역에 대한 컬러의 분포를 반영하는 단계별 값들을 추출한다. 표 3은 예로서 1, 3, 5 종류의 대표값 추출 방법이다.

4.3 평가

평가단계는 각 염색체에 대해 적합도를 계산하는 과정이다. 유전자 알고리즘에서 문제의 해를 결정하는 것은 바로 적합도 혹은 목적함수에 의해 결정된다. k번째 염색체에서 실제영상의 특성벡터 X와 염색체의 가상 벡터 Y에 대하여 R, G, B의 크기를 각각 R_{kx}, G_{kx}, B_{kx}와 R_{ky}, G_{ky}, B_{ky}로 나타내면 적합도는 다음과 같이 계산된다.

$$f(k) = -(|R_{kx} - R_{ky}| + |G_{kx} - G_{ky}| + |B_{kx} - B_{ky}|)$$

최적의 영상분할은 적합도 $f(k)$를 최대로 하는 영역들로 구성되었을 때로 정의되며, 분산 유전자 알고리즘은 진화 과정을 통해 최대 적합도를 갖는 최적의 염색체 군을 탐색한다.

4.4 선택 및 유전연산

선택단계에서 각 화소의 실제 RGB 값이 자기 자신을 포함해서 이웃(Neighborhood)의 크기에 따라 인접하는 염색체들 중 가장 좋은 적합도를 가지는 염색체로 대체되며, 이때 레이블도 대체된다. 이 과정에서 각 영역이 확대되거나 축소된다.

마지막으로 유전연산 단계에서는 우선 각 염색체에 대해 같은 레이블을 가지는 염색체와 일점 교차방법을 이용하여 교차한다. 교차하여 생성된 두 자식 염색체와 원래의 부모 염색체 중에서 적합도가 좋은 염색체로 대체하고, 이후 돌연변이를 하여 역시 적합도가 좋은 염색체로 대체한다. 돌연변이는 RGB 각각에 대해 그 값을 돌연변이율에 의해 무작위로 바꾼다. 교차율과 돌연변이율은 실험적으로 결정한다. 그림 3은 본 논문에서 제시하는 유전 연산의 예이다.

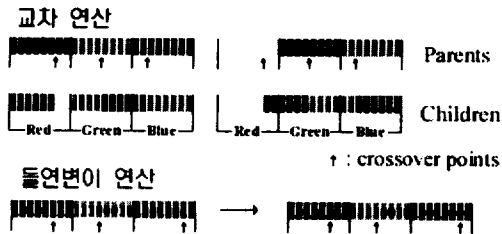



그림 3. 유전 연산(교차와 돌연변이)

이러한 과정을 종료 조건을 만족할 때까지 반복한다. 각 세대마다 이전 세대에 할당되었던 레이블이 새로운 세대에 얼마만큼 변하지 않고 안정되었나를 나타내는 안정도를 계산하여, 정해진 임계치를 넘어서는 횟수가 임계횟수이상 일어나거나 반복 세대수가 최대 반복횟수를 넘어서면 종료한다.

4.4 분할된 영역 통합

유방암조직영상에서 진단자가 관심을 가지는 의미있는 영역은 양성세포핵 영역, 음성세포핵 영역, 그리고, 배경 영역이며 크게 네 가지 색으로 나눌 수 있다. 염색체들을 초기화할 때 각 영역의 대표값들을 초기값으로 지정하면, 같은 영역에 대해서도 여러 개의 레이블 영역으로 분할된다. 따라서 레이블을 이용하여 하나의 영역으로 통합한다. 예를 들어 각 영역을 대표하는 세 종류의 대표값으로 초기화하면 염색체와 레이블은 12종류가 된다. 탐색과정이 끝나면 양성세포핵과 음성세포핵은 각각 세 종류, 배경은 여섯 종류의, 같은 영역이지만 서로 다른 레이블을 가진 영역들로 분할된다. 이러한 영역들을 초기화할 때 사용한 각 영역에 해당하는 레이블에 따라 하나의 영역으로 통합한다. 이러한 통합과정을 거쳐 최종적으로 양성세포핵 영역, 음성세포핵 영역, 그리고 배경 영역만으로 이루어진 결과 영상을 만든다.

5. 실험 및 분석

5.1 실험 방법

제안한 방법의 유효성을 검증하기 위하여 그림 4와 같이 펜티엄 PC, Windows 환경에서 Visual C++로 시스템을 구현하였다. 다양한 실험을 통하여 효과적인 분할 방법을 찾기 위해, 영역 성장을위한 비교시 이웃의 크기와 레이블 개수를 변화시키면서 실험하였다. 그리고 실제 영상에서 획득한 표본들에서 추출한 각 영역의 대표값으로 초기값을 지정하여, 초기값 임의 지정 방식과 비교 실험하였다. 종료조건으로 안정도 임계치는 99.5%, 임계횟수는 5회로 하였으며 최대 반복 횟수는 100회로 제한하였다. 표 4는 구체적인 실험 방법이다.

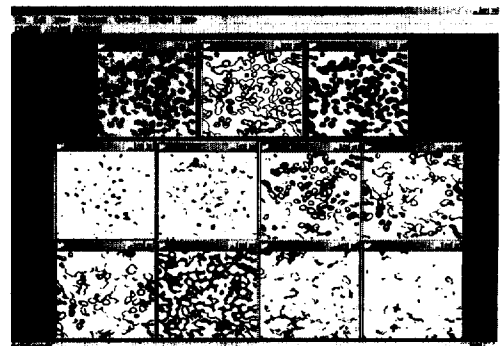



그림 4. 구현한 시스템

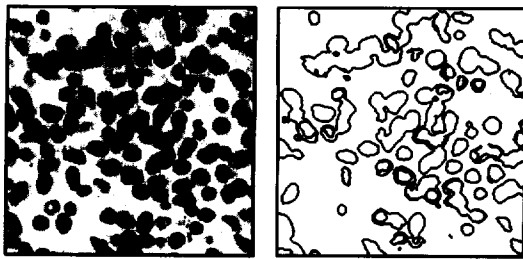
표 4. 실험 방법

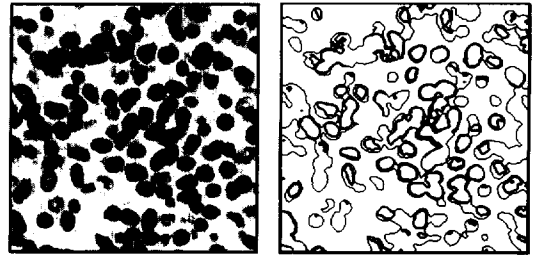
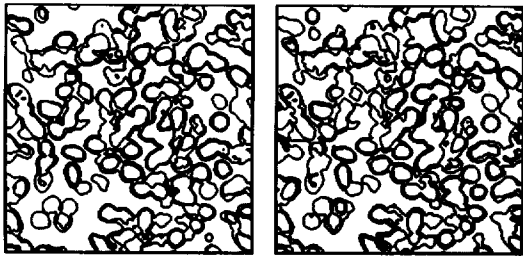
	실험방법1	실험방법2	실험방법3
이웃크기	3×3	3×3, 5×5, 7×7, 9×9, 11×11	3×3
초기값 지정방식	대표값	대표값의 평균	대표값
레이블개수	4,8,12,16	4	12,20,28

5.2 결과 분석

이웃의 크기는 3×3으로 고정한 상태에서, 초기값은 임의의 값으로 지정하고 레이블 개수를 4, 8, 12, 16으로 변화시킨 실험 방법 1의 결과는 그림 5와 같다. 레이블의 개수가 4일 때 가장 나쁜 결과를 보였으며, 동일한 영상일지라도 실험할 때마다 결과의 편차가 컸다. 그러나 레이블의 개수가 어느 정도 증가할수록 좋은 결과를 보였으며, 결과의 편차도 작았다. 특히 개수 12, 16일 때는 적절한 레이블 개수에 도달하여 차이를 보이지 않았다. 이는 레이블 수와 동일한 염색체의 종류가 증가할수록, 세포핵들을 대표하는 색과 가까운 초기값이 임의로 지정될 확률이 높기 때문이다. 그리고, 레이블 개수가 증가할 때, 사라지는 레이블도 상대적으로 증가하였다. 이는 유사한 염색체가 이웃에 존재할 때, 이웃 영역에 흡수되기 때문이다.

실험 방법 2에서는 이웃크기 변화에 따른 분할 결과를 분석하기 위해, 각 영역의 평균을 초기값으로 지정하고, 이웃크기를 3×3, 5×5, 7×7, 9×9, 11×11로 변화시키면서 실험하였다. 초기값을 임의로 지정하면 초기값에 큰 영향을 받기 때문에, 각 영역을 대표하는 컬러의 평균으로 초기값을 지정한 후, 이웃크기를 변화시키면서 실험하였다. 그림 6은 실험 방법 2의 결과이며, 이웃크기의 변화에 따른 차이는 크지 않다. 분산 유전자 알고리즘을 이용한 유방암조직영상 분할에서 이웃크기는 분할 결과에 거의 영향을 주지 않는다는 것을 알 수 있다.

그러나, 실험 방법 1에서 초기값 임의 지정 방식에 의한 네 종류의 염색체와 레이블을 사용하였을 때와,

(a) 테스트 영상

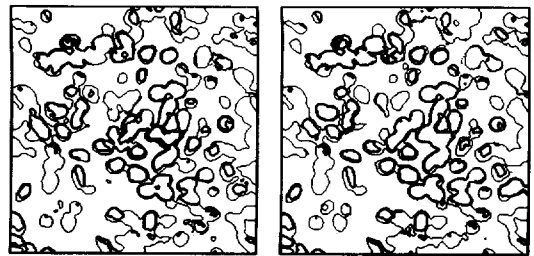
(b) label 개수 = 4

(c) label 개수 = 8

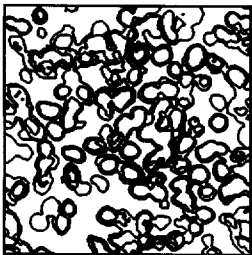
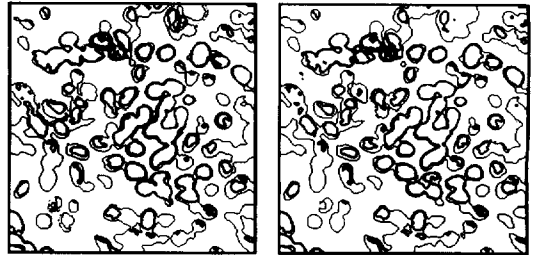
(d) label 개수 = 12

(e) label 개수 = 16

그림 5. 실험 방법1의 결과

(a) 테스트 영상

(b) 이웃크기 3×3

(c) 이웃크기 5×5

(d) 이웃크기 7×7

(e) 이웃크기 9×9

(f) 이웃크기 11×11

그림 6. 실험 방법 2의 결과

실험 방법 2에서 각 영역에서 추출한 대표값의 평균들을 초기값으로 지정한 네 종류의 염색체와 레이블을 사용하였을 때의 분할 결과는 큰 차이를 보였다. 이러한 결과는 성능을 향상시키기 위해서는 초기값으로서 각 영역을 대표하는 컬러를 사용해야 함을 제시한다.

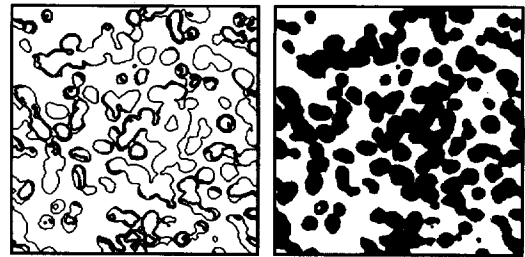
실험 방법 3에서는 양성세포핵의 갈색 영역, 음성세포핵의 검회색 영역, 그리고 배경을 이루고 있는 주황색과 노란색 영역에서 추출한 표본들의 평균, 최대, 최소 RGB값을 근거로, 각 영역에 대한 여러 단계의 대표값들을 추출하고 초기값으로 지정하였다. 표 5는 표본들로부터 계산된 각 영역에 대한 평균, 최대, 최소 값이며, 표 6은 이를 이용하여 추출한 각 영역에 대한 세 종류의 대표 값이다. 이웃 크기는 실험 방법 2의 결과에 따라 3×3으로 고정하였다. 그림 7은 표 6의 대표값으로 초기값을 지정했을 때의 결과로, (a)는 분할된 영역 사이를 선으로 구분한 결과이고, (b)는 시각적 효과를 위하여, 후처리로서 색상 처리한 결과이다. 파란색은 양성세포핵을 빨간색은 음성세포핵을, 그리고 흰색은 배경영역을 나타낸다.

다섯 종류, 일곱 종류의 다양한 대표값들을 초기값으로 지정하여 실험한 결과, 종류에 따른 분할 결과의 차이는 거의 없었다. 그러나 임의의 초기값과 대표값의 평균을 이용한 실험 방법 1, 2와는 차이를 보였다. 실험 방법 1의 동일 영상에 대한 편차 문제를 해결 할 수 있었으며, 실험 방법 2보다 좋은 결과를

표 5. 표본들의 평균, 최대, 최소값

데이터 종류		R	G	B
갈색 (양성)	최소	124	90	80
	평균	136	100	90
	최대	154	116	107
검회색 (음성)	최소	109	93	100
	평균	121	106	114
	최대	134	121	128
주황색 (배경)	최소	238	204	155
	평균	247	218	161
	최대	252	248	167
노란색 (배경)	최소	250	243	178
	평균	252	250	189
	최대	254	254	198

표 6. 영역당 세 종류의 대표값 추출 예

데이터 종류		R	G	B
갈색 (양성)	1	128	93	83
	2	136	100	90
	3	148	111	101
검회색 (음성)	1	113	97	104
	2	121	106	114
	3	130	116	124
주황색 (배경)	1	240	208	157
	2	247	218	161
	3	251	238	165
노란색 (배경)	1	250	245	181
	2	252	250	189
	3	254	253	195

(a) 선 구분 결과

(b) 색 구분 결과

그림 7. 실험 방법 3의 결과

보였다. 이는 각 영역에 대한 다 단계 대표값을 적용함으로써 양성세포핵와 음성세포핵의 중간 단계의 화소에 대한 변별력을 높였기 때문이다. 그림 8은 실험 방법 3으로 여러 종류의 유방암조직영상을 분할했을 때의 결과이다.

6. 결 론

본 논문에서는 분산 유전자 알고리즘을 이용하여 컬러 유방암조직영상을 분할하는 방법을 제안하였다. 기존의 명암값 영상에 대한 방법을 개선하여, 컬러 정보를 효과적으로 반영하는 염색체, 평가함수 및 유전연산을 고안하여 탐색 능력을 향상시켰다. 또한 다양한 실험을 통하여 이웃 크기, 레이블 개수, 초기값의 상관관계를 분석하였다.

실험에서 일반적인 방법과 같이 초기값을 임의로

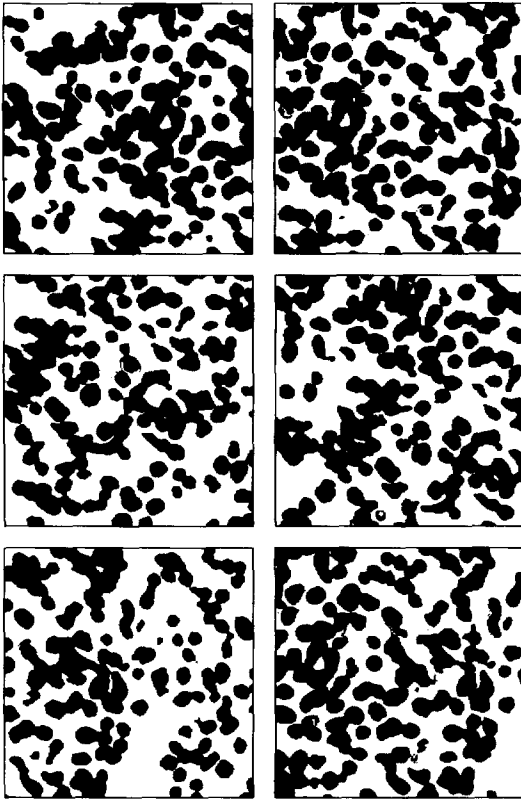



그림 8. 다양한 영상에 대한 실험 방법 3의 결과

지정했을 때, 레이블 개수가 증가할수록 좋은 결과를 보였다. 염색체의 종류가 증가할수록, 세포핵들을 대표하는 색이 임의로 지정될 확률이 높기 때문이었다. 또한 이웃의 크기에 따른 차이는 거의 보이지 않았다. 따라서 분할 결과에 큰 영향을 미치는 초기값으로, 전처리 과정을 통해 추출된 대표값들을 이용하였을 때, 효과적인 분할이 이루어 졌다. 더불어 의료영상의 분석에서 중요한 요소인 재현가능 문제를 해결할 수 있었다.

결론적으로 유방암조직영상과 같이 획득하고자 하는 개체들이 정해져 있고 컬러의 종류도 뚜렷할 때, 영역 기반의 자율 탐색 방법인 분산 유전자 알고리즘은 좋은 분할 결과를 보였다. 또한 적절한 초기값 지정은 중간 단계의 색깔을 가진 모호한 개체를 효과적으로 분류하여 성능을 향상시켰다.

향후 연구과제는 유방암조직영상과 유사한 다양한 의료영상에 대한 응용이며, 분야별 초기값 추출방법론에 대한 것이다.

참 고 문 헌

[1] P. Ranefall, "Towards Automatic Quantification of Immunohistochemistry Using Colour Image Analysis," Ph. D. Thesis, Uppsala University, Sweden, ISBN 91-554-4152-1. pp. 12-14, 1998.

[2] D. H. Ballard and C. M. Brown, *Computer Vision*, p. 116, Prentice-Hall, Inc., EngleWood Cliffs. NJ. 1982.

[3] H. K. Choi, T. Jarkrans et al., "Image Analysis Based Grading of Bladder Carcinoma. Comparison of Object, Texture and Graph Based Methods and Their Reproducibility," *Analytical Cellular Pathology*. Vol.15, pp. 1-18, 1997.

[4] 최현주, 허민권, 최흥국, 김상균, 최항묵, 박세명, "칼라 유방암조직영상에서 질감 특징과 신경회로망을 이용한 양성세포핵과 음성세포핵의 자동분할," 한국정보과학회 학술발표논문집, 제26권, 제2호, pp. 422-424, 1999.

[5] 김재륜, 하진영, 김백섭, 김호성, "자궁경부암 세포 영상 분할을 위한 Thresholding 기법," 한국정보과학회 학술발표논문집, 제26권, 제2호, pp. 419-421, 1999.

[6] Tetsuo Asano. Naokazu Yokoya, "Image Segmentation Schema for Low-level Computer Vision," *Pattern Recognition*, Vol.14, pp. 267-273, 1981.

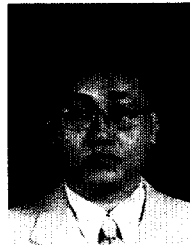
[7] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. PAMI*, Vol.6, No.6, pp. 721-741, 1995.

[8] D. K. Panjwani and G. Healey, "Markov Random Field Models for Unsupervised Segmentation of textured Color Images," *IEEE Trans. PAMI*, Vol.17, No.10, pp. 939-954, 1995.

[9] 이경미, 김상균, 김항준, "유전자 알고리즘을 이용한 MRF기반의 Texture 분할," 한국정보처리학회 논문지, 제5권, 제10호, pp. 713-2724, 1998.

[10] Philippe Andrey and Philippe Tarroux, "Unsupervised Image Segmentation using a Distributed Genetic Algorithm," *Pattern Recognition*, Vol.27, No.5, pp. 659-673, 1994.

조 찬 윤

1999년 인제대학교 전산학과(이학사)
1999년~현재 인제대학교 전산학과 석사과정 재학중
관심분야 : 컴퓨터비전, 패턴인식

김 상 균

1991년 경북대학교 통계학과(이학사)
1994년 경북대학교 대학원 컴퓨터공학과(공학석사)
1996년 경북대학교 대학원 컴퓨터공학과(공학박사)
1996년~현재 인제대학교 정보컴퓨터공학부 조교수
관심분야 : 정보처리